

## 連続音声認識エキスパートシステムとその支援環境

## — 音声認識と知識工学の融合 —

辻野 克彦 溝口 理一郎 角所 収

(大阪大学産業科学研究所)

1. まえがき

筆者等は、「音響分析結果の時間変化を視察することによってかなりの範囲で音韻を決定できる音声研究者は、視覚による音声認識のエキスパートである。」という考えに基づき、このエキスパートの振舞を知識工学的手法を用いてシミュレートする音声認識システムの試作を行ってきた。[1]

本研究においては、連続音声認識のための知識ベースを構築することが極めて重要な役割を占めていることは言うまでもない。しかしながら現時点におけるシステムは、知識ベースの管理機構及び構築支援ツール等は整備されておらず、知識ベースを構築する環境としては問題があった。開発初期においてはルールの数も少なく開発者自身によるルールの管理は容易であったし、修正の際の支援ツールの必要性は感じられなかった。ところがルール数が100を越え、システムの性能がある程度高くなるにつれて開発環境の整備が重要な問題となってきた。即ち、現在の環境の下ではこれ以上の性能向上を行うことにかなり大きな困難が伴うようになってきた。現在我々は、これらの問題を解決するために知識ベース構築支援環境の整備を進めている。考察の対象としては、

- (1) ルールの構成を見直し、ルール配置等に含まれる隠れた情報を明白に表現する。
  - (2) ルールの管理・検索・検証各機能の開発。
  - (3) 音声データベースとその関連づけ。
  - (4) ルール作成支援環境の開発。
  - (5) マンマシンインターフェースの改善。
- 等があげられる。

近年、知識ベース管理に関する研究が盛んに行われているが、タスクに依存しない一般的な議論が多いように見受けられる。[2], [3] 一方、我々の立場は連続音声の認識という特定の領域における知識ベースに関するものである。知識ベースの研究においては、一般的理論と同様に特定の領域における具体例を通じた経験に基づく考察が極めて重要であると考えられている。その意味で、我々が現在のシステムの開発を通して得た経験をもとにとして知識ベース管理の問題を考察することは、単に一つの知識ベースを開発することに留まらず、本格的な知識ベース研究のためのアプローチとして位置付けられる。

本稿の概要は以下の通りである。

2章では従来のシステムの概要について述べる。3章では知識ベースを管理し開発してゆく上での問題点について述べ、

4章に各問題点に対する解決策を述べることによって、本システムの基本仕様を明らかにする。5章ではLISPマシン(Symbolics 3600)上でのインプリメントの現状と、設計されたシステムの動作の概要を示す。

## 2. 従来のシステムの概要

システムは、入力となる連続音声の音響的処理を行う特徴抽出部および、音韻認識を行う記号処理部からなり、記号処理部においてプロダクションルール表現された知識を用いて処理が行われるものであった。セグメンテーション、子音認識、母音認識にエキスパートの知識のルール表現を用いるが、従来の認識システムに比べて、

- (1) システムの性能向上が、常にルールの修正あるいは追加という手段で容易に行える。
- (2) システムの実行手順(記号表現された時系列のデータと用意された知識とのマッチング操作)とシステムの作成者の認識過程が同じであるため、現在不足する知識、ルールが必要とする条件等の検出が容易である。
- (3) 認識に有用な特徴パラメータの追加が、プロダクションシステムにおけるワーキングメモリーの増加とそれを用いるルールの追加にどどまるため、システムは拡張性に富む。
- (4) 調音結合を音韻の再認識処理の立場から積極的に利用することができ、音韻認識率を向上させることができる。

といった利点を持つ。人間のエキスパートの視覚による音韻認識率(子音はいくつのグループ単位に分類される)は、言語情報を考慮しないで、約95%であることが確かめられており、これが本システムの到達目標となっている。システムの試作段階で、成人男性3名の発声した約30秒の連続音声に関して、セグメンテーションの正答率93%、音韻認識率85%なる結果が得られている。

本システムの処理の流れの概要を以下に示す。(図1参照)

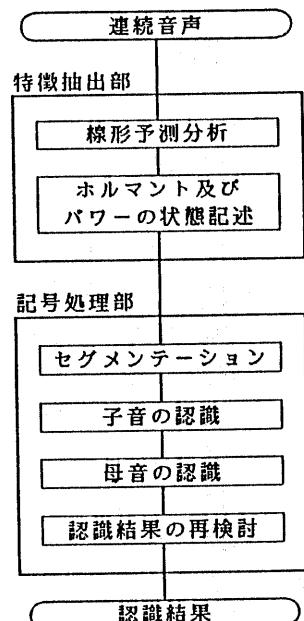


図1 処理手順の概略

### 特徴抽出部

- (1) パワー、ホルマント周波数、バンド幅、ピッチ周波数、及び、パワーの高域成分比の時間変化を分析する。(図2)
- (2) ホルマント周波数及びバンド幅か

らなる2次元平面においてクラスタリングを行いホルマントを抽出する。

- (3) 根の変動の傾向を考慮して、滑らかな変動を起こしている部分では連続性を重視するが、特異な変動は見落とさないように各パラメータ時間変化を記号化する。(図3)

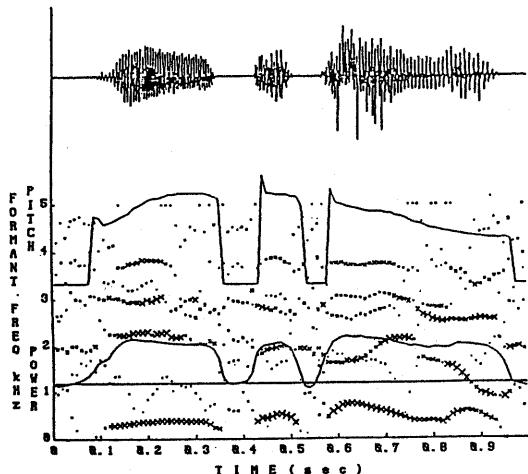


図2 連続音声の分析結果

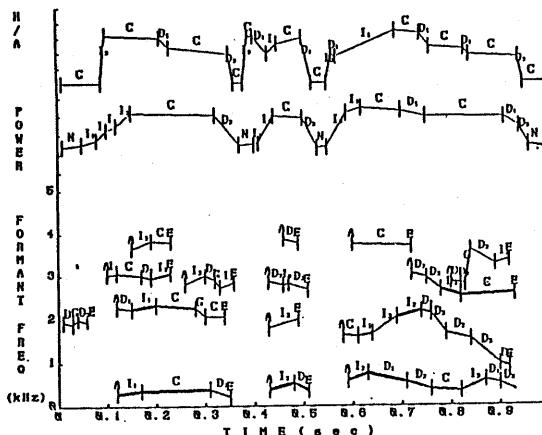


図3 分析結果の記号表現

#### 記号処理部

- (4) 音韻境界に見られるパワーのDIPやホルマントの出現・消滅・不連続点に着目し、セグメント境界を検出する。
- (5) 特徴が明白な子音の認識を行う。
- (6) 子音と認識されなかった部分について母音認識を行う。
- (7) /OUA/と認識されてしまった/OWA/や欠落した/KYO/の中の/Y/などの音韻は、前後のセグメントの認識結果や調音結合を考慮して正しく認識し直す。

特徴抽出部はスーパーミニコンMV/800Ⅱ上のFORTRANで、記号処理部は大型計算機上のLISPで書かれたプロダクションシステムHIPS[4]を用いて実現されている。

#### 3. 記号処理部の問題点

現在、記号処理部はセグメンテーション用として47個、子音認識用として29個、母音認識用として15個、再認識用として21個、計112個のルールからなっている。ルールの総数の増加に加えて発火条件の複雑なものも増え、これによって、以下に示すような3点が大きな問題となってきた。

- (1) 現システムにおいては、ルールはソースコードのなかで物理的に上に書かれたものが優先するようになっており、この順序づけに含まれる意味、すなわちルール間の階層関係が隠されてしまう。このため後々のルール修正によって、この関係が乱された場合、システムはうまく作動しなくなってしまう。

さらに、新しくルールを追加する場合でも、それを付け加える場所によってルールの動作が保証できなくなることが考えられる。

- (2) ルールの修正の際、そのルールがいつ誰によってなぜ作られたのか、作成の根拠となった音声試料はどれであったか、というような情報が失われているため、そのルールの正当性や有効性が把握できず、修正した結果、もどもとの機能を失ってしまうことがある。
- (3) ルール及び音声試料が増えて来たため、ルールの改良・開発・検証の際参照する音声試料や関連するルールを探し出し、利用しやすく整理するのが容易でなくなってきたこと。

#### 4. 支援環境の設計

前章で述べた問題点を解決するため、システム全体の見直しと知識ベース管理機構及びルール開発支援環境の整備が必要となった。図4に支援環境を含めたシステムの全体構成を示す。また、以下に各部分に必要な機能と役割を示し、前章の問題点の解決案について述べる。

##### 4.1 ルールデータベース

後で述べるシステムインタフェースと協力して、ルールの管理、その正当性のチェックなどを行う。ルールを管理するための情報としては、

- ルール名、作成者、日時
- ルールが結論する音韻カテゴリ
- ルールのドキュメンテーション
- ルール作成の目的ならびに基準となった音声試料
- ルールが参照する特徴パラメータ
- 関係する他のルール
- そのルールによって認識された音声データベース中の音韻などが必要であり、これらは検索の際のキーとして使われる。

ルールの検索に関しては、あるルールの持つ属性を求める向きの検索と、ある属性をもつルールを求める向きの検索が考えられる。

##### 4.2 音声データベース

システムインタフェースに協力して、ルールの開発・改良・検証の際に必要な音声試料を管理し、要求された音韻部分の音声波形や状態記述を提供する。これにより、

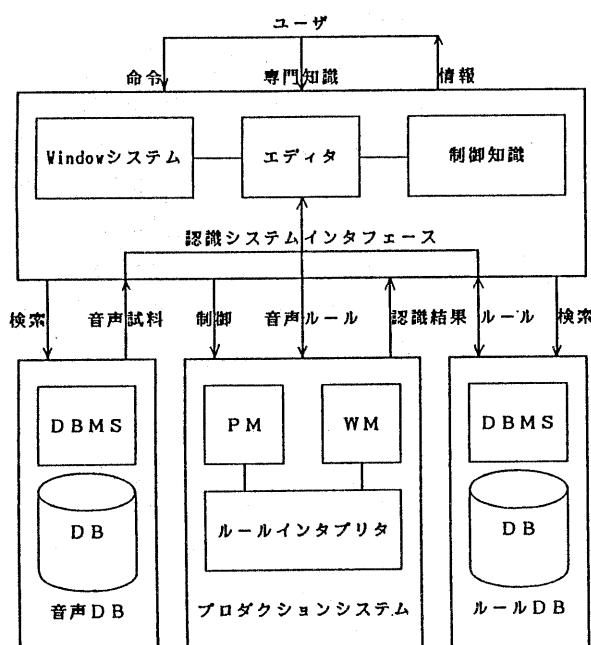


図4 システム構成

認識実験やルールの検証の際の音声試料集めに要する労力が大幅に軽減される。

音声データベースは音声データを音声ファイルに対するポインタとして管理している。[5] 音声データは、音韻・発声者などをキーとして検索ができる。

#### 4.3 認識部

認識部はシステムインターフェースから与えられた音声データとルールを用いて認識をおこない結果を返す。

新しい認識部は Lisp Machine Symbolics 3600上の高速プロダクションシステム OPS 5eで実現される。認識状況やルール適用の様子は Window に表示され、システムの状況の把握やデバッグをしやすくする。

#### 4.4 認識システムインターフェース

この部分は、ユーザとシステム(ルールデータベース、音声データベース、認

コマンド名	機能	引数
find	データベース検索要求 結果をテンポラリバッファ (以下TB)に記録	1 検索する要素数 2 "data" or "rule" 3 ~ :key 要素
print	TBデータの表示	1 ~ :key
load	TBデータの読み込み	1 バッファ名
use	PM・WMの設定	1 :PM バッファ名 or :WM バッファ名
xy-plot	TBデータのグラフ表示	1 x 軸 :key 2 y 軸 :key
start	認識開始 引数キー情報をTBに記録	1 ~ :key

図5 コマンドの概略

識部)との間にあってユーザ支援の要となる。つまり、ユーザの要求に応じて、音声に関する知識や統計知識を用いて2つのデータベースの情報を収集・整理してユーザに示したり、それを認識部に渡して認識実験をしたりする。ルール修正・開発の際、ルールデータベースと協力して新しいルールの正当性等をチェックする機能をもつ。

このインターフェースに対するコマンドの例とその機能を図5に示す。

#### 4.5 問題点の解決

3章で述べた問題点は以上に述べた支援環境を相互に関連付けることにより解決できる。

ルール構造に依存する問題点(1)を解決するために、ルールに Context, KS (Knowledge Source)等の階層的な構造を導入し、ルール全体を新しいプロダクションシステム上に再構築する。KSは、他のKSをコンテキストによって手続き的に活性化することができる。たとえば、図6に示した音韻/MN/を認識するためのKSは1つのトップレベルルールと/MN/判定用の下位ルール群から成っている。トップレベルルールの条件部は、KSのプリコンディションに対応する。トップレベルルールはその実行部で母音中心判定KSを起動して中心フレームを決定させ、下位ルールにその結果を用いて認識を行わせる。

ルール管理に関する問題点(2)は、ルールのドキュメンテーションやそのルールの作成の根拠とな

ルール名	/MN/トップレベル
所属KS	/MN/認識
コンテキスト	'子音認識
条件部	UV 有声音
	PWR ベグメント内で減少
	F1 セグメント境界でギャップ
実行部	母音中心判定KS発動
	下位ルール用評価基準計算
	コンテキスト変更 'MN認識
	( /MN/ 判定用下位ルール群起動 )
ルール名	/MN/下位レベル-1
所属KS	/MN/認識
コンテキスト	'MN認識'
条件部	F2 > 2600Hz
実行部	認識結果 = /MN/
	コンテキスト変更 '再認識'
ルール名	/MN/下位レベル-X
所属KS	/MN/認識
コンテキスト	'MN認識'
条件部	他の下位レベルルールがマッチしない
実行部	認識結果 = 不明
	コンテキスト復活 子音認識

図6 /MN/ルールの例

った音声試料についての情報をルールデータベースに持たせて、認識システムインターフェースがアクセスできるようにすることにより解決する。これはルールデータベースにおいてはルールから属性の検索をすることに対応する。

支援環境は問題点(3)において最も効力を發揮する。たとえば、誤認識が起きたことがユーザによって知らされたとき、認識システムインターフェースはルールデータベースと音声データベースを効率的に結びつけて、その部分の音韻とそれに関連するルールや音韻を表示し、ユーザが直ちに検討することができるよう手配する。又、修正されたルールを検証するための関連する音声試料やルール検索も行う。この場合ルールデータベースは属性からルールの検索を行うことになる。

## 5. システム操作の様子

現在設計中のシステムの予想される動作の例を図7に示す。これは、認識実験中に誤認識を発見したときの対応の例である。

- (1) ユーザは、認識実験中に誤認識を発見する。
- (2) そこでユーザはプログラムの実行を中断させて、デバッグすることにした。この際音声データの詳細が別ウインドウに表示される。
- (3) ルールデータベースに命じて、/MN/認識用のルールのなかで、H/Aパラメータを参照しているものを1つ探ってきて、そのルールのドキュメントを画面に表示させた。
- (4) ここで、ユーザは H/Aパラメータの閾値が悪いのであろうと考えて、その様子を調べる実験をするため /MN/ と /I/ 認識ルールのすべてと 音韻 /M/・/N/・/I/ を検索して、それぞれバッファに書き出した。
- (5) それぞれのバッファをプロダクションメモリとワーキングメモリに割り当てて、
- (6) プロダクションメモリをエディタで編集して閾値を変更した。エディタwindowは必要などきにのみ現われる。
- (7) 検証実験をおこなった。

この例では、コマンド入力や結果の出力等、ほとんどすべての表示が一画面上で行われているが、Lisp Machineの特徴を生かして、メニュー入力・ウインドウ出力等を採用すればより使い良いシステムとなると考えられる。

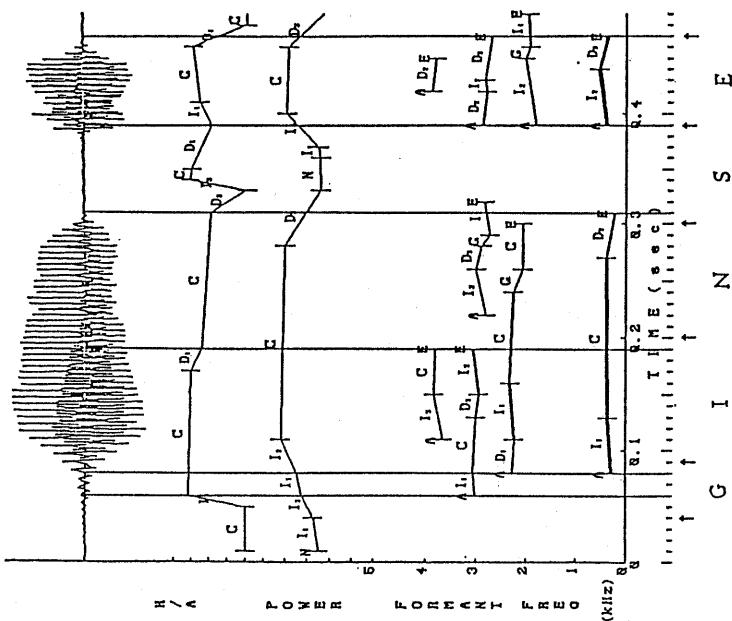
start	Segment	Phoneme	Rec			
1	G	BDG	-			
2	I	I	-			
3	N	I	-			
4	S	SH	-			
5	E	[Abort]	-			
User Abort	... [Abort	[Debug	? ]			
Segment Start frame	Rules	F1	F2	H/R	Phone me	Rec
2	38	BOIN-3	366	2319	62	I
3	54	BOIN-B-2	366	2068	43	N
						I

**RULE MN-1**  
 $\sigma_{MN}$  is very similar to  $\sigma_{II}$ , but it differs in the point below.  
 F1 is flat in the segment,  
 but it has Gap or Rapid change near the segment-border.  
 PHR is also flat in the segment,  
 and the level is lower than each side vowel segment.  
 And if  
 HRF parameter is less than 25(?) , or F2 is greater than 2600Hz,  
 it's MN.

```
find all data :phonene M N I ;load buffer-1  
547 data found.  
Data loaded into Buffer-1  
  
find all rule :kind MN I ;load buffer-2  
22 rule(s) found.  
Rules loaded into Buffer-2
```

```
use :PM buffer-2 :WM buffer-1  
edit PM  
Rule M-1 modified 8: {H/A <25.} ==> {H/A <45.}
```

start :F1 :F2 :H/A :output @LGP		F1	F2	H/A	Phonee	Rec
1	358	2323	44	I	NN	
2	424	2544	25	NN	NN	
3	361	1967	66	I	I	
4	388	2244	62	I	I	
5	351	1665	59	I	I	



#### 図7 予想される動作例

卷之三

198 L'ECOLE DE

```

>find all data :phoneme MN1 ;load buffer-1
346 data found.
Data loaded into Buffer-1

>find all rule :kind MN1 ;load buffer-2
20 rule(s) found.
Rules loaded into Buffer-2

>use :PM buffer-2 :WM buffer-1

>start :H/A :F2 ;xy-plot :H/A :F2

```

図8 グラフ表示コマンドの例

また図8のようなユーザの要求に従って各種情報をグラフ等に整理して表示する機能や統計的あるいは数学的処理を施す機能もインプリメントする。これにより、本章第1節で述べた H/Aパラメータの閾値決定の際に、/M/, /N/, /I/ 分布の様子をグラフに表示できるので、閾値の決定が非常に容易になる。またこの場合などは、閾値をシステムに自動的に決定させることも可能である。

以上に述べた支援の多くはユーザからの要求に従い情報を提供する受動的なものである。更に進んだ知識ベース支援環境として、統計情報やユーザへの質問を通した知識獲得などの能動的支援を考えられるが、これは将来の課題したい。

## 6. むすび

知識工学的手法を用いた連続音声認識システムの支援環境について述べた。

現在、認識部、認識ルールの再構成と支援環境の外部仕様、論理設計を進めている。今後これらのシステムをLisp Machine上にインプリメントして支援環境の整備を行う予定である。

## (参考文献)

- [1]福田、溝口 他：“知識工学的手法を用いた連続音声認識”，信学技報，EA84-66 (1985).
- [2]北上、国藤 他：“大規模な知識ベース管理システムをめざして”，情報処理学会知識工学と人工知能研究会資料36-7 (1984).
- [3]新谷：“知識ベースにおける知識記憶管理機構の試作”，情報処理学会第30回全国大会，2L-8 (1985).
- [4]溝口、富田 他：“階層的プロダクションシステム HIPS”，情報処理学会論文誌，Vol.23, No.2, PP.177-186 (1982).
- [5]溝口、前田 他：“知的アクセス機能をもつ音声データベース「S P E C H - D B 」”，情報処理学会論文誌，Vol.24, No.3, PP.271-280 (1983).